

Evolution of Molybdenum Nitrogenase during the Transition from Anaerobic to Aerobic Metabolism

Eric S. Boyd,^a Amaya M. Garcia Costas,^b Trinity L. Hamilton,^b Florence Mus,^b John W. Peters^b

Department of Microbiology and Immunology, Montana State University, Bozeman, Montana, USA^a; Department of Chemistry and Biochemistry, Montana State University, Bozeman, Montana, USA^b

ABSTRACT

Molybdenum nitrogenase (Nif), which catalyzes the reduction of dinitrogen to ammonium, has modulated the availability of fixed nitrogen in the biosphere since early in Earth's history. Phylogenetic evidence indicates that oxygen (O₂)-sensitive Nif emerged in an anaerobic archaeon and later diversified into an aerobic bacterium. Aerobic bacteria that fix N₂ have adapted a number of strategies to protect Nif from inactivation by O₂, including spatial and temporal segregation of Nif from O₂ and respiratory consumption of O₂. Here we report the complement of Nif-encoding genes in 189 diazotrophic genomes. We show that the evolution of Nif during the transition from anaerobic to aerobic metabolism was accompanied by both gene recruitment and loss, resulting in a substantial increase in the number of *nif* genes. While the observed increase in the number of *nif* genes and their phylogenetic distribution are strongly correlated with adaptation to utilize O₂ in metabolism, the increase is not correlated with any of the known O₂ protection mechanisms. Rather, gene recruitment appears to have been in response to selective pressure to optimize Nif synthesis to meet fixed N demands associated with aerobic productivity and to more efficiently regulate Nif under oxic conditions that favor protein turnover. Consistent with this hypothesis, the transition of Nif from anoxic to oxic environments is associated with a shift from posttranslational regulation in anaerobes to transcriptional regulation in obligate aerobes and facultative anaerobes. Given that fixed nitrogen typically limits ecosystem productivity, our observations further underscore the dynamic interplay between the evolution of Earth's oxygen, nitrogen, and carbon biogeochemical cycles.

IMPORTANCE

Molybdenum nitrogenase (Nif), which catalyzes the reduction of dinitrogen to ammonium, has modulated the availability of fixed nitrogen in the biosphere since early in Earth's history. Nif emerged in an anaerobe and later diversified into aerobes. Here we show that the transition of Nif from anaerobic to aerobic metabolism was accompanied by both gene recruitment and gene loss, resulting in a substantial increase in the number of *nif* genes. While the observed increase in the number of *nif* genes is strongly correlated with adaptation to utilize O₂ in metabolism, the increase is not correlated with any of the known O₂ protective mechanisms. Rather, gene recruitment was likely a response to more efficiently regulate Nif under oxic conditions that favor protein turnover.

All life requires fixed nitrogen (N), and its availability often limits ecosystem productivity (1, 2). Most of the N on Earth is in the form of dinitrogen (N₂), which is unreactive, bio-unavailable, and must be chemically reduced to ammonium (NH₄⁺) before it can be incorporated into biological molecules, such as proteins or nucleic acids. The primary enzyme that catalyzes the reduction of N₂ to NH₄⁺ is the molybdenum (Mo)-dependent nitrogenase Nif (3, 4). Our recent phylogenetic studies indicated that Nif emerged in methanogens (5, 6), implying an origin in an anoxic and possibly sulfidic environment. Given the ecological advantage afforded to populations capable of fixing N₂ and the metabolic expense of maintaining the genetic machinery necessary to synthesize active Nif (Fig. 1A), selection would presumably direct biology of organisms to evolve mechanisms to protect Nif from oxidative damage and which would allow diazotrophs to diversify into oxic niches. Indeed, Nif has been identified in a diversity of microorganisms, including obligate aerobes and oxygenic phototrophs (5, 7, 8), that have evolved several physiological strategies that allow diazotrophic growth in oxic environments (9–12). These strategies include (i) temporal decoupling where organisms fix N₂ at night when O₂ tensions are lowered due to colocalized aerobic respiration activity (*Synechococcus*-like cyanobacteria). (ii) spatial decoupling through the segregation of nitrogenase to

anaerobic heterocyst organelles in an otherwise-oxygenated environment (*Anabaena*-like cyanobacteria), and (iii) metabolic decoupling by which organisms effectively maintain an anoxic cytoplasm through increased metabolic consumption of O₂ (*Azotobacter*-like *Proteobacteria*).

Nif consists of a homodimeric NifH component (Fe protein) that donates electrons to the heterotetrameric NifDK component (MoFe protein), which contains the O₂-sensitive iron-molybdenum cofactor (FeMo-co), the site of substrate reduction (4). In addition to these nitrogenase structural proteins, *nif* gene clusters

Received 23 December 2014 Accepted 24 February 2015

Accepted manuscript posted online 2 March 2015

Citation Boyd ES, Costas AMG, Hamilton TL, Mus F, Peters JW. 2015. Evolution of molybdenum nitrogenase during the transition from anaerobic to aerobic metabolism. *J Bacteriol* 197:1690–1699. doi:10.1128/JB.02611-14.

Editor: W. W. Metcalf

Address correspondence to John W. Peters, john.peters@chemistry.montana.edu.

Supplemental material for this article may be found at <http://dx.doi.org/10.1128/JB.02611-14>.

Copyright © 2015, American Society for Microbiology. All Rights Reserved. doi:10.1128/JB.02611-14

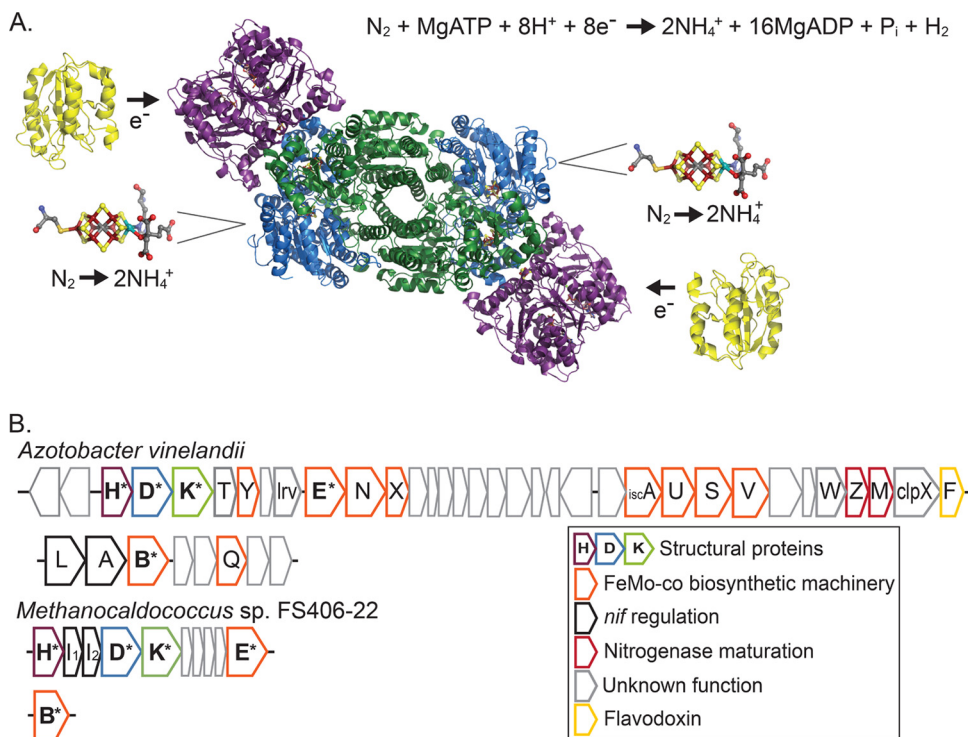


FIG 1 (A) Schematic of the reduction of N_2 to NH_4^+ by Nif, with the reaction stoichiometry indicated. Electrons carried by flavodoxin (yellow) are transferred to NifH (purple), which interacts with and transfers electrons to the P cluster, located between NifD (blue) and NifK (green), and ultimately to the FeMo-co cluster of NifD, where N_2 reduction occurs. (B) Structure and composition of *nif* gene clusters in *Azotobacter vinelandii* AVoP (both major and minor *nif* clusters) and *Methanocaldococcus* sp. FS406-22. Asterisks and boldfaced letters delineate the minimal complement of genes required to form an active nitrogenase (*nifHDKEB*).

can encode proteins involved in regulation (NifALI₁I₂), electron transfer (NifFJ), and FeMo-co biosynthesis (NifXENBQUVYS), as well as proteins with as-yet-undefined functions (13). The functional composition of *nif* gene clusters varies considerably among taxa (5, 8, 14), where they range from the minimum gene complement known to encode an active nitrogenase (*nifHDKEB*) as well as several regulatory and uncharacterized genes (e.g., *Methanocaldococcus* sp. strain FS406-22) to those comprising >20 genes (e.g., *Azotobacter vinelandii*) (Fig. 1B). The observation of simplified *nif* gene complements in anaerobic methanogens and more extensive suites of genes in obligate aerobes, coupled with previous work indicating that Nif emerged in a hydrogenotrophic methanogen in an anoxic environment (5, 6), suggests that the number of *nif* genes has increased over evolutionary time.

Two ancillary *nif*-associated complexes (Rnf and Fix) have been proposed to function in electron transfer to Nif (see Table 2, below). The Rnf complex (*R*hodobacter *n*itrogen *f*ixation) was first identified in the diazotroph *Rhodobacter capsulatus*, where it was postulated to donate electrons to Nif (20) but has since been shown to assume a number of alternative roles in other diazotrophs as diverse as being essential for early transcription of *nif* genes in *Azotobacter vinelandii* (55) to involvement in NH_4^+ -mediated “fast switch off” of Nif in *Azoarcus* sp. strain BH72 (56). The Fix complex has also been implicated in electron transport to Nif (57). FixAB is related to electron-transferring flavoprotein (Etf), and FixCX is related to Etf-quinone reductase; it has been proposed that FixABCX functions in some aerobic diazotrophs to bifurcate electrons from NADH to ferredoxin and ubiquinone

(21). In this reaction the coupling of the endergonic reduction of ferredoxin by NADH is driven by the accompanied exergonic oxidation of NADH by ubiquinone as a component of the respiratory chain. This allows a proportion of the electron flux from NADH to be directed to Nif.

Here, we examine the composition of *nif* genes in 189 diazotrophic organisms in reference to the evolutionary history of Nif and the metabolic strategies for reducing damage to nitrogenase by O_2 . The results indicate that O_2 had a profound influence on the evolution of Nif through the recruitment of specific accessory proteins that are involved in enzyme regulation and maturation. Surprisingly, the recruitment of these *nif* genes appears to have been independent of the emergence of metabolic strategies for reducing damage to nitrogenase by O_2 . Rather, results suggest that recruitment of *nif* genes was driven by selective pressure to optimize Nif synthesis to meet fixed N demands associated with increased productivity in aerobes relative to anaerobes and to more efficiently regulate Nif under oxic conditions that favor protein turnover. Consistent with this hypothesis, our data indicate that the diversification of Nif during the transition of diazotrophs from anoxic to oxic environments was accompanied by a shift from posttranslational regulation of Nif in anaerobes to transcriptional regulation in obligate aerobes and facultative anaerobes. Such observations further underscore the dynamic and complex interplay between the evolution of Earth’s oxygen, nitrogen, and carbon biogeochemical cycles.

MATERIALS AND METHODS

Diazotrophic taxa and their complement of *nif* genes. We defined the minimum gene complement required to synthesize active Nif as consisting of *nifHDKBE*, based on two recent studies that indicated N₂ fixation in organisms with this minimal gene complement (15, 16). Moreover, the active site protein environment and dimensions of nitrogenase from organisms having a minimal gene complement were recently examined and shown to likely harbor FeMo-co (17). Nif protein homologs were compiled from all completed and publicly available genome sequences by using BLASTp in conjunction with the Department of Energy Integrated Microbial Genomes (DOE-IMG) server in January 2012. NifHDKBE proteins from *A. vinelandii* AvOP served as query sequences. This search revealed a total of 189 taxa that harbored homologs of NifHDKBE (see Table S1 in the supplemental material). Alignment and screening for cluster binding residues that demarcate these proteins were performed as previously described (5). Despite the presence of cluster binding residues in these 189 taxa, it is important to note that only a fraction of the 189 organisms identified as having the minimal *nif* gene complement have been biochemically or physiologically shown to be capable of reducing N₂.

The *nifHDK* and *nifEB* homologs are generally clustered on the chromosome, although they can be present in separate gene clusters (18). The composition and order of genes flanking homologs of *nifHDKBE* were determined manually by using the Neighborhood viewer on the DOE-IMG server. Genome sequences were screened for the presence of homologs of *nifNXZYQVSUFMWALI₂OT* and *nafY*, *iscA* *lrv*, and *clpX* (associated with *nif*), and manual assignments were verified by using reciprocal BLASTp (see Table S1 in the supplemental material). The composition of gene clusters associated with alternative nitrogenases, which have yet to be identified in a genome that does not also encode *nif* (5, 6), were not included in phylogenetic and evolutionary analyses, since they appear to utilize components of the *nif* system to synthesize their active site metal clusters (19). The FixABCX protein sequences (encoded by *fixABCX*) from *A. vinelandii* were used as BLASTp queries in order to determine if the 189 genomes of taxa harboring homologs of *nifHDKBE* also encoded Fix. Organisms with *fixABC* or *fixABCX* were considered positive for encoding the Fix system. For Rnf, we used the deduced amino acid sequence from the *rnfABCDGEH* genes in *Pseudomonas stutzeri* as the query for a BLASTp search. Genomes with 4 or more of those genes were considered to encode an Rnf system.

Phylogenetic analysis. Individual H, D, and K homologs were aligned as described previously (6) with ChlLNB/BchLNB from *Anabaena variabilis* ATCC 29413 and *Chlorobium limicola* DSM 245 serving as outgroups. The individual alignment blocks were concatenated with PAUP (version 4.0) (22) and subjected to evolutionary model prediction with ProtTest (version 2.4) (23). The phylogeny of each concatenated protein sequence was evaluated with PhyML-aBayes (version 3.0.1) (24), specifying the LG amino acid substitution matrix with a discrete 4 category gamma substitution model (gamma shape parameter, 0.872) and a defined proportion of invariant sites of 0.032. Approximate likelihood ratio tests (aLRT) were used as an alternative to nonparametric bootstrap frequencies. A consensus phylogenetic tree was projected from 1,000 aLRT permutations using FigTree (version 1.2.2; <http://tree.bio.ed.ac.uk/software/figtree/>). The phylogram was rate smoothed by using a penalized likelihood approach (25) with the chronopl program, specifying a lambda smoothing parameter of 1.0 over 1,000 iterations. The chronopl program is part of the Ape package (version 3.0-3) (26) and is implemented with the base package R (version 2.13.1) (27).

Individual genes were treated as binary traits (see Table S2 in the supplemental material), and trait-based evolutionary methods were applied to determine the extent to which NifHDK phylogeny predicts the distribution of *nif* genes among taxa. The phylogenetic signal (K-statistic) associated with the dispersion of individual *nif* genes on the rate-smoothed NifHDK chronogram was quantified using the program multiphylo within the Picante package (28) as implemented with the base package R. The K statistic compares the observed signal in the distribution of a trait (e.g., an individual *nif* gene) on a phylogeny to the signal under a Brownian motion model of evolution (29). Values of K that are close to 1 imply a Brownian motion for the evolution of that trait (or some degree of phylogenetic signal), while values greater than 1 indicate a strong phylogenetic signal. K values closer to zero or which are negative correspond to a random or convergent pattern of evolution for that trait. The statistical significance of phylogenetic signals was evaluated by comparing patterns in the variance of independent contrasts of the trait to a null model produced by shuffling taxa labels across the tips of the phylogeny. The distribution of individual genes among diazotrophic taxa was evaluated by mapping their dispersion on the NifHDK phylogenetic tree by using the Ape package.

The number of *nif* genes in a diazotroph (*nif* gene content) was calculated by dividing the sum of the number of *nif* gene homologs present by 21, the total number of *nif*-associated genes examined in the present study. The relationship between the presence or absence of a given gene (treated as a binary data set) and the nature of the metabolism of the organism (binary data set for aerobicity; 1 denotes an aerobe or facultative anaerobe and 0 denotes an anaerobe) was evaluated using Pearson linear regression within the program XL Stat (version 2008.7.03). A matrix describing the Jaccard dissimilarity in the *nif* genes in diazotrophs (treated as a binary data set) was generated using base functions in R. The matrix was analyzed using principal coordinates (PCO) analysis in order to visualize patterns in the dissimilarity in the composition of the *nif* genes. The complexity of *nif* genes was overlaid (i.e., “surfing”) on the surface of the ordination by using the R programs LabDSV (<http://ecology.msu.montana.edu/labdsV/R/>) and Vegan (<http://CRAN.R-project.org/package=vegan>).

The mechanisms that aerobic or facultative anaerobic organisms use to protect Nif from oxidative damage were characterized and used to evaluate the relationships between these mechanisms and individual *nif* genes (see Table S2 in the supplemental material). The protective mechanisms were classified as temporal segregation, heterocyst formation (spatial segregation), metabolic or respiratory protection, or the formation of bacteroids or vesicles (spatial segregation). A value of 1 was assigned to taxa that utilize a particular mechanism, whereas a value of 0 was assigned to taxa that do not use that particular mechanism. Variation in this binary data set in relation to the presence/absence of particular *nif* genes was evaluated using linear regression approaches described above.

RESULTS AND DISCUSSION

Number of *nif* genes as a function of taxonomy and Nif phylogeny. We analyzed *nif* genes from aerobic, facultative anaerobic, and anaerobic N₂-fixing microorganisms. Among the 3,182 total genomes (121 archaeal and 3,062 bacterial) publicly available from DOE-IMG in January 2012 (see Table S1 in the supplemental material), a total of 189 were identified that had *nif* genes (see Table S1), roughly 5.9% of the sequenced prokaryotes. The *nif* genes identified in these taxa were distributed within a single archaeal phylum and among 12 bacterial phyla (Fig. 2). The number of Nif-associated genes in these taxa ranged from a minimum of 7 to a maximum of 20 genes, with an average of 13 genes. The most simple suites of *nif* genes were identified in the genomes of hyperthermophilic methanogens (e.g., *Methanocaldococcus* spp.), which consist of *nifHDKBEI₂*, while the most complex suites were identified among members of the *Gammaproteobacteria* (e.g., *A. vinelandii* and *P. stutzeri*), which encode *nifHDKENBXAZVSUFWTZOML* as well as *nafY*, *iscA*, *lrv*, and *clpX*, which colocalize with the *nif* genes. The putative roles for the products of some of the genes that vary in their distribution among genomes are shown in Table 1.

The number of *nif* genes in each diazotrophic genome was used to calculate a metric describing the “complexity” of each suite of *nif* genes, where a value of 0 indicated a *nif* gene content with

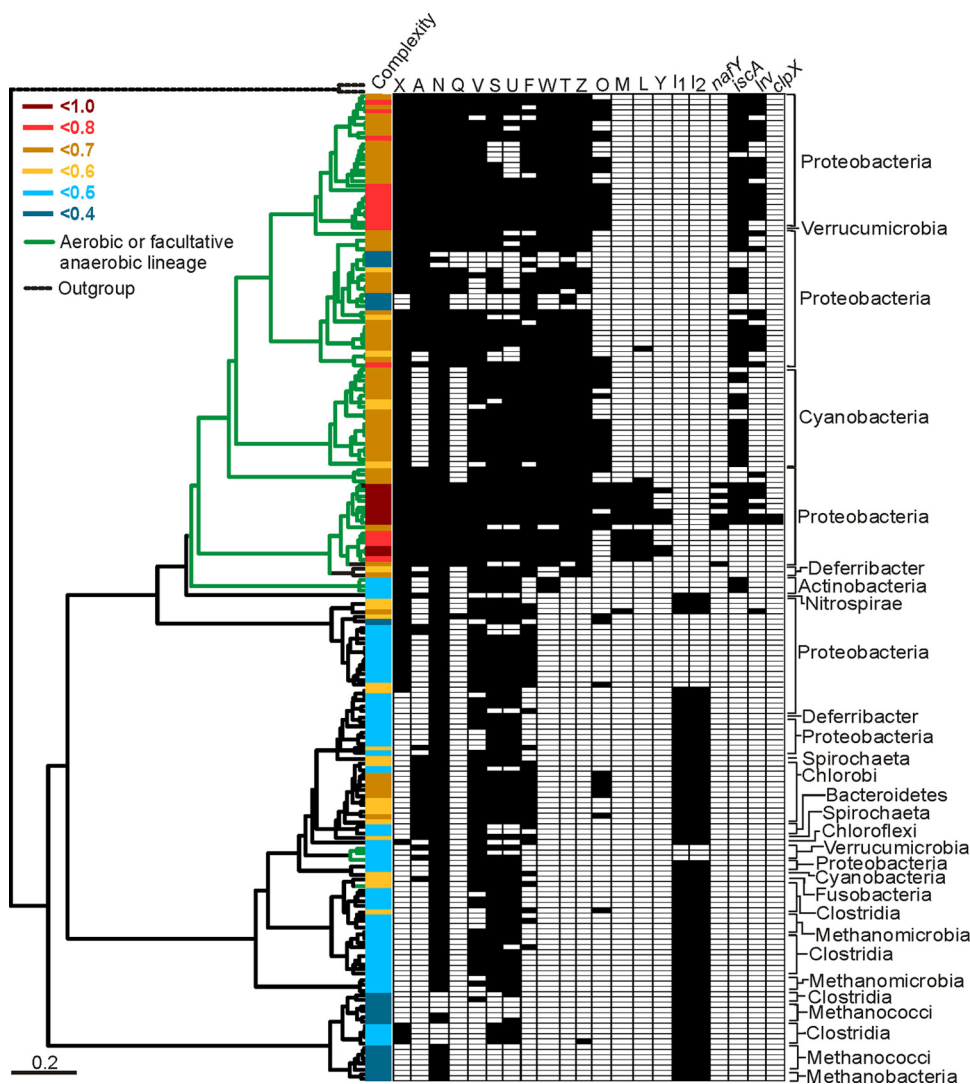


FIG 2 NifHDK chronogram from 189 taxa, with an overlay of aerobic/facultative anaerobic lineages in green. The outgroup lineage for the rooted tree is depicted by a dotted line. The terminals of lineages are depicted by a heat map indicating the complexity of the *nif* gene clusters. The presence (black boxes) or absence (white boxes) of individual *nif* gene cluster-associated loci as well as the phylum-level classification for each taxon are indicated. See Table S1 in the supplemental material for additional details of the *nif* gene contents and gene accession numbers for each taxon.

minimal complexity (few *nif* genes) and a value of 1.0 indicated maximal complexity (many *nif* genes). The complexity of *nif* gene content ranged from 0.26 to 0.91, corresponding to methanogen and gammaproteobacterial genomes, respectively. To examine relationships between the complexity of *nif* gene content and the evolutionary history of Nif, the complexity metric was overlaid on a concatenated NifHDK phylogenetic tree as a heat map (Fig. 2, side bar). Qualitatively, this analysis indicated that the complexity of *nif* gene content increased substantially during the diversification of organisms that encode Nif, with early evolving lineages (e.g., methanogens, clostridia) tending to harbor less complex suites of *nif* genes than more recently evolved lineages (e.g., *Proteobacteria*, *Cyanobacteria*).

To identify the progression in the recruitment of *nif* genes, individual *nif* genes for each taxon were mapped on the NifHDK phylogenetic tree (Fig. 2), and the covariation in the distribution of *nif* genes was evaluated by using linear regression approaches

(Table 2). All of the 21 genes examined exhibited significantly positive correlations ($P < 0.05$) with the complexity of *nif* gene content. In particular, pronounced associations were identified between the complexity of *nif* gene content and *nifZ* (Pearson R , 0.80; $P < 0.001$), *nifW* (Pearson R , 0.78; $P < 0.001$), *nifT* (Pearson R , 0.75; $P < 0.001$), and *nifQ* (Pearson R , 0.73; $P < 0.001$). The limited evidence available on the putative roles of these gene products suggests that they may have been recruited to improve FeMo-co or MoFe protein maturation and/or to enhance the stability of these metal clusters. Specifically, NifZ has been proposed to function in P cluster biosynthesis in the MoFe proteins of *A. vinelandii* (30, 31) and *Klebsiella pneumoniae* (32) and may be involved in the repair of this cluster. *A. vinelandii* and *K. pneumoniae* Δ *nifZ* mutant strains still have measurable, albeit reduced, N_2 reducing activities relative to the corresponding wild types, indicating that NifZ is not essential (30–32). Likewise, *A. vinelandii* and *K. pneumoniae* Δ *nifW* mutants exhibit reduced nitroge-

TABLE 1 Proposed functional roles of genes identified in *nif* gene clusters

Gene	Proposed function	Reference(s)
<i>nifW</i>	MoFe protein interaction; may contribute to homocitrate processing	35
<i>nifQ</i>	Molybdenum donor to FeMo-co	36
<i>nifT</i>	Unknown	40
<i>nifX</i>	Transient reservoir of FeMo-co	62
<i>nifZ</i>	Facilitates P cluster assembly	31
<i>nifA</i>	Positive regulator of nitrogen fixation	50, 64
<i>nifF</i>	Flavodoxin; donates electrons to NifH in <i>K. pneumoniae</i> and maybe in <i>A. vinelandii</i>	65–67
<i>nifJ</i>	Electron transport to FeMo-co	65, 68
<i>iscA</i>	Putative role in maturation of 4Fe-4S clusters under aerobic growth or oxidative stress conditions	69
<i>nifM</i>	Stabilizes NifH	33, 70
<i>nifL</i>	Oxygen sensor and negative regulator of nitrogen fixation	51, 53, 71, 72
<i>nifY</i>	Stabilizes apodinitrogenase and facilitates FeMo-co insertion	73
<i>nafY</i>	Stabilizes apodinitrogenase and facilitates FeMo-co insertion	74
<i>nifV</i>	Homocitrate synthase	13, 75
<i>nifS</i>	Cysteine desulfurase (Fe-S clusters)	13, 76
<i>nifU</i>	Molecular scaffold for assembly of Fe-S clusters	13
<i>nifH1</i>	Posttranslational regulation of nitrogen fixation in anaerobes	41–43
<i>nifH2</i>	Posttranslational regulation of nitrogen fixation in anaerobes	41–43

nase activity compared to the corresponding wild types, suggesting that the gene product, although also not essential, enhances FeMo-co biosynthesis and enzyme activity (32–34). Intriguingly, NifW forms a complex with the MoFe protein when exposed to O₂, indicating that NifW might be involved in protecting Nif from

inactivation by O₂ (35), although a specific role for NifW in O₂ protection has not been ascribed. While *nifZ* and *nifW* are nonessential and appear to have been recruited to improve protein maturation and/or enhance enzyme stability, *nifQ* is required for Mo-nitrogenase activity in *A. vinelandii* (36) and *K. pneumoniae* (37)

TABLE 2 Correspondence (Pearson *R* values) between the presence/absence of individual *nif* genes, general physiological characteristics of diazotrophic organisms containing *nif*, and mechanisms of protection of Nif from inactivation by O₂

Variable or <i>nif</i> gene	Correspondence (<i>R</i>) between gene and ^a :		O ₂ protection mechanism				
	Complexity	Aerobic/facultative anaerobic metabolism	Temporal	Heterocyst	Metabolic	Bacteroids	Vesicles
Complexity	1.00	0.62	0.12	0.06	0.19	−0.08	−0.08
Aerobic condition	0.62	NA ^b	NA	NA	NA	NA	NA
<i>nifX</i>	0.59	0.64	0.15	0.14	0.12	0.11	0.10
<i>nifY</i>	0.40	0.12	−0.05	−0.03	0.18	−0.05	−0.02
<i>nafY</i>	0.36	0.14	−0.05	−0.04	0.17	−0.06	−0.03
<i>nifQ</i>	0.73	0.63	−0.18	−0.13	0.21	0.16	−0.09
<i>nifV</i>	0.55	0.21	0.14	0.10	0.08	−0.35	0.07
<i>nifS</i>	0.45	0.09	0.12	0.01	0.07	−0.20	0.06
<i>nifU</i>	0.32	−0.07	0.14	0.11	0.00	−0.37	0.07
<i>nifF</i>	0.59	0.40	0.19	−0.05	0.11	0.06	−0.17
<i>nifM</i>	0.56	0.22	−0.07	−0.05	0.09	−0.09	−0.04
<i>nifW</i>	0.78	0.85	0.23	0.20	0.16	0.08	0.14
<i>nifT</i>	0.75	0.84	0.22	0.20	0.16	0.20	−0.12
<i>nifZ</i>	0.80	0.80	0.22	0.20	0.16	0.08	−0.12
<i>nifA</i>	0.56	0.45	−0.25	−0.18	0.15	0.30	−0.13
<i>nifL</i>	0.57	0.26	−0.07	−0.05	0.09	−0.09	−0.04
<i>nifH1</i>	−0.58	−0.79	−0.16	−0.15	−0.12	−0.24	−0.11
<i>nifH2</i>	−0.58	−0.79	−0.16	−0.15	−0.12	−0.24	−0.11
<i>nifO</i>	0.55	0.38	0.25	0.15	−0.01	−0.05	−0.08
<i>iscA_{nif}</i>	0.64	0.73	0.19	0.05	0.19	0.10	0.17
<i>lrn</i>	0.57	0.45	−0.12	−0.09	0.11	−0.05	−0.06
<i>clpX</i>	0.27	0.11	−0.03	−0.02	0.34	−0.03	−0.01
<i>nifN</i>	0.31	0.14	0.06	0.04	0.04	−0.10	0.03
<i>rnf</i> operon ^c	0.35	0.02	−0.15	−0.11	0.16	−0.18	−0.08
<i>fix</i> operon ^c	0.14	0.29	−0.15	−0.11	0.08	0.41	−0.08

^a Values in boldface showed significant correspondence ($P < 0.05$).

^b NA, not applicable.

^c Not necessarily located in or near *nif* gene cluster.

TABLE 3 Phylogenetic signals (*K* statistics) associated with distribution of individual *nif* genes or general physiological traits when mapped on the NifHDK cladogram^a

Trait or gene	<i>K</i> statistic	Statistical significance ^b (observed vs random) based on:	
		<i>P</i> value	Z-score
<i>nifI</i> ₁	4.18	0.001	-1.70
<i>nifI</i> ₂	4.18	0.001	-1.65
Aerobic conditions	1.37	0.001	-1.73
<i>nifW</i>	1.36	0.001	-1.76
<i>nifZ</i>	1.26	0.001	-1.72
<i>nifQ</i>	0.97	0.001	-1.52
<i>nifX</i>	0.88	0.001	-1.61
<i>nifA</i>	0.60	0.001	-1.67
<i>nifL</i>	0.46	0.001	-0.72
<i>nifM</i>	0.45	0.001	-0.78
<i>iscA</i> _{nif}	0.33	0.001	-1.65
<i>nifT</i>	0.25	0.001	-1.80
<i>nifV</i>	0.21	0.001	-1.34
<i>nifS</i>	0.18	0.001	-1.26
<i>lrv</i>	0.16	0.001	-1.18
<i>nifN</i>	0.14	0.002	-0.66
<i>nifO</i>	0.11	0.001	-1.40
<i>nafY</i>	0.09	0.003	-0.49
<i>clpX</i> _{nif}	0.08	0.030	-0.25
<i>nifY</i>	0.08	0.004	-0.44
<i>nifU</i>	0.02	0.053	-1.36
<i>nifF</i>	0.01	0.070	-1.68

^a *K* values (for the phylogenetically independent contrasts) near 0 indicate a random pattern of evolution for a given trait, whereas values of 1 or greater indicate a strong phylogenetic signal for a given trait.

^b The statistical significance levels of a *K* statistic (*P* value and Z score) were evaluated by comparing variations in the observed distribution of a gene on the NifHDK tree with that generated by randomly shuffling taxa labels.

when grown under ambient conditions, where it functions in Mo acquisition for FeMo-co biosynthesis.

However, when supplied with excess molybdate (1,000-fold) in the growth medium, $\Delta nifQ$ strains demonstrated Nif activity similar to wild type (37, 38). It is possible that *nifQ* was recruited to meet Mo requirements and to maximize the efficiency of FeMo-co biosynthesis in order to meet fixed N needs in organisms operating high-energy-yielding metabolisms. Of all of the *nif* genes that appear to have been recruited in response to a transition from anaerobic to aerobic/facultatively anaerobic metabolism, *nifT* is arguably the least well characterized. *nifT* was shown to be nonessential in *A. vinelandii* (39), and overexpression of *nifT* in *K. pneumoniae* did not affect regulation, maturation, or activity of Nif in the presence of ammonium or under N₂-fixing conditions (40). The only two genes whose distribution exhibited a negative correlation with the complexity of *nif* gene content were *nifI*₁ and *nifI*₂ (Pearson *R*, -0.58; *P* < 0.001), both of which function in the posttranslational regulation of Nif in response to the availability of fixed sources of N (41–43). These genes are only present in strictly anaerobic organisms (see Table S1 in the supplemental material).

All *nif* genes exhibited a positive phylogenetic signal (*K* statistic) when mapped on the NifHDK phylogeny (Table 3). The genes with the highest *K* values were *nifI*₁/*nifI*₂ (*K* = 4.18; *P* = 0.001), *nifW* (*K* = 1.36; *P* = 0.001), and *nifZ* (*K* = 1.26; *P* = 0.001). These observations suggest a general trend of recruitment and vertical

inheritance of *nif* genes, although gene loss (e.g., *nifI*₁ and *nifI*₂) and horizontal gene transfer events have also impacted the evolution of Nif (3, 7, 8, 44).

Influence of O₂ on the evolution of Nif. We defined the overall metabolic strategy of taxa that include *nif* genes as either aerobic (including facultative anaerobic) or anaerobic, and we overlaid this on the NifHDK phylogeny in order to evaluate the relationship between metabolism and (i) overall *nif* gene content and (ii) the phylogenetic distribution of individual *nif* genes. A clear demarcation in the evolutionary history of NifHDK was detected between organisms harboring strict anaerobic metabolism and aerobic metabolism (Fig. 2). NifHDK is associated with strictly anaerobic taxa branched at the base of the tree, which is consistent with previous analyses that indicate Nif evolved in an anaerobe and in an anoxic environment (5, 6). In contrast, NifHDK is associated with aerobic/facultative anaerobic taxa from several recently evolved lineages. The ability to use O₂ in metabolism was significantly and positively correlated with the complexity of *nif* gene content (Pearson *R*, 0.62; *P* < 0.001) (Table 2). Moreover, when treated as a binary trait, the ability to use O₂ in metabolism had a significant phylogenetic signal (*K* = 1.37; *P* < 0.01) (Table 2) when mapped on the NifHDK phylogeny, indicating a strong pattern of vertical inheritance for the physiological capacity to integrate O₂ into metabolism with respect to Nif evolution. This may be due to a strong selective pressure to synthesize a more effective Nif that is capable of taking advantage of additional reductant and ATP made available by high-energy-yielding aerobic metabolisms in order to meet the fixed N demands of more-productive cells.

Principal coordinates (PCO) analysis was used to examine the extent to which the overall complement of *nif* genes associated with each taxon reflected the ability to utilize O₂ in metabolism (Fig. 3). PCO analysis revealed a clear demarcation in the complement of *nif* genes of taxa that employ anaerobic metabolism, compared with those that utilize aerobic/facultative anaerobic metabolism. PCO axis 1 (50.2% of variance explained) was significantly correlated with the ability to utilize O₂ in metabolism (Pearson *R*, 0.87; *P* < 0.01) and with the complexity of the complement of *nif* genes associated with each taxon (Pearson *R*, 0.76; *P* < 0.01). PCO axis 2 (10.0% of variance explained) was inversely correlated with

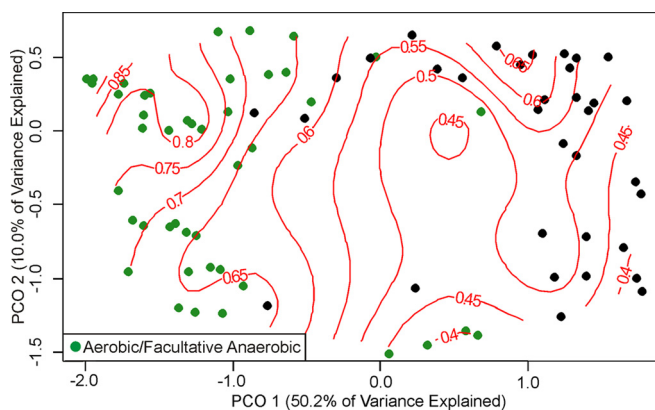


FIG 3 Principal coordinates analysis results for a matrix describing the Jaccard dissimilarity of 189 *nif* gene clusters. For simplicity, complements of *nif* genes that were identical in composition are depicted by a single dot. The calculated complexity of *nif* gene contents, as plotted on the surface of the PCO ordination, yielded a *D*² of 0.95, indicating a near-perfect fit. *nif* gene complements from aerobic or facultative anaerobic taxa are indicated by green dots.

the complexity of *nif* gene content (Pearson R , 0.18; P = 0.02) but was not correlated with the ability to utilize O_2 in metabolism. The complexity of *nif* gene content was plotted on the surface of the PCO ordination, yielding a D^2 of 0.95, which indicates a near-perfect fit of the model to the data. *nif* gene complexity isopleths were oriented primarily along PCO axis 1, consistent with the separation of *nif* gene content in organisms operating with aerobic versus anaerobic metabolisms along this axis. The results indicated a significant correspondence between (i) the ability to utilize O_2 in metabolism, (ii) the composition of *nif* genes associated with these taxa, and (iii) the complexity in *nif* gene content among taxa.

Linear regression analyses were used to identify genes that were likely recruited in response to the adaptation to utilize O_2 in metabolism (Table 2). Positive correspondences were noted between the ability to utilize O_2 in metabolism and the taxonomic distribution of all *nif* genes, with the exceptions of *nifl*₁ and *nifl*₂, which both exhibited strong and inverse correlations with the use of O_2 in metabolism (Pearson R , -0.79). The strongest correspondences were noted between the ability to utilize O_2 in metabolism and the distribution of *nifW* (Pearson R , 0.85), *nifT* (Pearson R , 0.84), *nifZ* (Pearson R , 0.80), and *iscA_{nif}* (Pearson R , 0.73). The basis for the positive association between the distribution of *nifT* (function unknown) and *nifZ* (involved in P cluster assembly) and the ability to integrate O_2 into metabolism is not clear. In contrast and as described above, NifW has been shown to form a complex with the MoFe protein when exposed to O_2 and thus may play a role in protecting Nif from inactivation by O_2 in aerobes (35), although a specific role for NifW in O_2 protection has not been ascribed. The distribution of NafY also exhibited a positive, albeit weak, correlation (Pearson R , 0.14) with the ability to use O_2 in metabolism. NafY, which is not required for diazotrophic growth of *A. vinelandii* under normal conditions (30°C), is required when grown diazotrophically under temperature stress at 37°C (45). This suggests that NafY may have a stabilizing effect on the MoFe protein or function to mitigate damage to FeMo-co clusters under stress conditions. *iscA_{nif}* is thought to provide an alternative scaffold for mediating *nif*-specific and general Fe-S cluster assembly (46). It is plausible that *iscA_{nif}* was recruited to better coordinate the Fe-S cluster assembly demands of Nif and other housekeeping functions. Collectively, these results suggest that the ability to utilize O_2 in metabolism has been a strong selective pressure driving the recruitment (*nifW*, *nifT*, *nifZ*, and *iscA_{nif}*) or loss (*nifl*₁ and *nifl*₂) of genes from *nif* gene clusters during the evolution of Nif.

The relationship between mechanisms of O_2 protection and *nif* gene content. Nif is irreversibly inactivated by O_2 (13). As such, diazotrophs have evolved mechanisms that spatially or temporally separate N_2 fixation from aerobic respiration or which render the Nif environment anoxic (respiratory protection) (9–12). We evaluated whether each protective mechanism is associated with recruitment of Nif-specific genes by using linear regression (Table 2). Metabolic or respiratory protection of Nif was the only mechanism that correlated, albeit weakly (Pearson R , 0.19), with *nif* gene content. Moreover, we found no strong correlation between any of the aforementioned mechanisms and the distribution of a particular gene in the associated *nif* gene clusters (Table 2). The strongest correlation identified was between *clpX* and metabolic or respiratory protection of Nif (Pearson R , 0.34), which is proposed to be important for aerobic diazotrophy in *A. vinelandii* (10). The formation of bacteroids is also weakly associated with the distribution of *nifA* (Pearson R , 0.30) and inversely associated

with the distribution of *nifU* (Pearson R , -0.37) and *nifV* (Pearson R , -0.34). None of the heterocyst- or bacteroid-producing microorganisms or those microorganisms that employ a temporal strategy to protect Nif from O_2 has genes that typify “complex” *nif* gene complements, in particular, those that encode NifMLY and NafY. In contrast, all of the genomes of bacteroid-producing microorganisms contain *fix* genes (described below). Overall, the lack of strong correlations between O_2 protective mechanisms and the distribution of individual *nif* genes suggests that the selective pressure to recruit additional *nif* genes was different from the selective pressure that drove the adaptation of O_2 protection mechanisms.

Nif regulation during the transition from anaerobic to aerobic metabolism. Nif is tightly regulated due to the metabolic costs associated with fixing N_2 (minimally, 16 mol ATP per mol reduced) and with maintaining the repertoire of genes required to synthesize active Nif (Fig. 1). Several *nif*-encoded proteins are involved in regulation, including Nifl₁I₂, NifA, and NifL (12, 47, 48). While the distributions of *nifl*₁ and *nifl*₂ in gene clusters were strongly and inversely correlated with the ability to use O_2 in metabolism (Pearson R , -0.79), the distributions of *nifA* and *nifL* exhibited positive correlations with the ability to use O_2 in metabolism (Pearson R , 0.42 and 0.26, respectively). This suggests a shift in the primary means by which Nif is regulated during the transition from anaerobic (Nifl₁I₂) to aerobic (NifAL) metabolism. Nifl₁ and Nifl₂ are paralogous and are both members of the P_{II} family of nitrogen-regulatory proteins (49). These proteins form a complex with Nif, suppressing N_2 reduction at the posttranslational level (41, 43). Posttranslational suppression of Nif is relieved by 2-oxoglutarate (2-OG) (42). The concentration of 2-OG increases during N starvation and eventually reaches a critical concentration that interrupts the Nifl₁I₂ interaction with Nif-HDK, thereby allowing N_2 fixation to resume (47). *nifA* is widely distributed in aerobes (the primary exception being cyanobacteria) but is uncommon in anaerobes (the primary exception being *Chlorobi*), where it functions as a transcriptional activator in combination with the RNA polymerase sigma factor RpoN (50). NifA activity is prevented by O_2 and ammonium (51) and also is affected by 2-OG concentrations (52). However, in some *Gamma-roteobacteria*, the negative effector NifL is required to prevent NifA-mediated activation of *nif* genes in response to O_2 or NH_4^+ availability (12, 51, 53). These *Gammaproteobacteria* are among the earliest evolving facultative anaerobes or obligate aerobes (Fig. 2), indicating that the NifAL mechanism of regulation might have played an important early role in adapting Nif to oxic conditions, with *nifL* being lost in more-recently evolved lineages.

Several lineages of organisms encode both Nifl₁I₂ and NifA regulatory mechanisms (anaerobic *Chlorobi*), while other lineages (e.g., aerobic cyanobacteria and anaerobic *Geobacter* spp.) do not encode either of these regulatory mechanisms. Nif regulation in cyanobacteria is complex and likely involves at least transcriptional-level control (9), whereas Nif regulation in *Geobacter* spp. is thought to be under the control of a histidine-aspartate phosphorelay system under the control of a RhoN-dependent promoter (54).

Relationships between metabolism, *nif* gene content, and the distributions of *fix* and *rnf* gene loci. Because respiration and reduction of dinitrogen to ammonium must compete for electrons under low reduction potentials, we examined whether there was a correlation between two *nif*-associated complexes (Rnf and

Fix) that have been proposed to function in electron transfer to Nif (Table 2). Organisms that encode the Rnf system tend to contain more complex complements of *nif* genes (Pearson *R*, 0.35) (Table 2). Intriguingly, a strong association between aerobic lifestyle, which is strongly correlated with the complexity of *nif* gene content (Pearson *R*, 0.62), and the presence of the *rnf* genes was not observed (Pearson *R*, 0.02). This is due to the fact that not all aerobes encode *rnf* and some of those taxa that do encode *rnf* also have an intermediate number of *nif* genes. In contrast, the presence of *fix* genes was correlated with the ability to utilize O₂ in metabolism (Pearson *R*, 0.29). This suggests a potential relationship between the need to balance the flux of electrons going to Nif and those going to generate ATP through respiratory processes. This refinement might be due to the requirement for both reducing equivalents and ATP for Nif and the requirement for reducing equivalents to generate ATP through respiration. It has been proposed that FixABCX functions in some aerobic diazotrophs to bifurcate electrons from NADH to ferredoxin and ubiquinone (21). In this reaction, the coupling of the endergonic reduction of ferredoxin by NADH is driven by the accompanied exergonic oxidation of NADH by ubiquinone as a component of the respiratory chain. This allows a proportion of the electron flux from NADH to be directed to Nif. Interestingly, the association between the distribution of *fix* genes and diazotrophs that form bacteroids is also significant and positive (Pearson *R*, 0.41); the reason for this observation is not clear.

Acquisition and inheritance of *nif* genes. It is intriguing to speculate about the origin of the *nif* genes that appear to have been recruited in aerobes. The distribution of genes that track strongly with the ability to utilize O₂ in metabolism (i.e., *nifW* and *nifX*) in recently evolved anaerobic lineages suggests that the products of these genes could have played key roles in the transition of Nif from anoxic to oxic conditions. *nifW* is found in representatives of all aerobic diazotrophic lineages, including the *Actinobacteria* (see Table S1 in the supplemental material), a group identified as harboring the earliest-evolving Nif proteins among aerobic/facultative anaerobes and the group in which *nifW* was first identified (Fig. 2). *Actinobacteria* segregate Nif to vesicles (58), where it is protected by an external hopanoid lipid envelope that limits O₂ diffusion (59). Hopanoids are produced by bacteria under low-O₂ conditions (60), and it is possible that NifW, which has been proposed to form a complex with the MoFe protein when exposed to O₂ (35), facilitated the early acquisition and function of Nif in these taxa. *nifX* is also present in all aerobic genomes and in some anaerobic proteobacterial genomes (Fig. 2), suggesting recruitment of *nifX* took place prior to the transition from anoxic to oxic N₂ fixation. NifX binds FeMo-co or FeMo-co precursors, and while it is not required for Mo-dependent N₂ fixation (61, 62), it has been speculated to increase the efficiency of FeMo-co biosynthesis by providing protection to the O₂-labile FeMo-co and its intermediates (13). It is interesting that anaerobic organisms and facultative anaerobes or aerobes that lack an independent *nifX* gene encode a NifB with a fused NifX domain (*nifB-nifX*) (5). This suggests that the fusion event resulting in the *nifB-nifX* protein may be uncoupled to the selective pressure associated with aerobic diazotrophy. *iscA_{nif}*, although present in representatives of all aerobic phyla, appears to have been lost from many genomes (Fig. 2). This may be due to the presence of functionally redundant *iscA* genes in those taxa and a biological impetus (i.e., selection) to purge genomes of unused or redundant functionalities in competi-

tive ecological situations (63). *nifT* and *nifZ* are present in all lineages with the exception of *Actinobacteria*, suggesting that they were recruited to *nif* gene clusters after the divergence of this lineage. For unknown reasons, *nifA* and *nifQ* are absent in the *Actinobacteria* and the cyanobacteria. Intriguingly, nearly all organisms that lack *nifH₁I₂* encode NifA, a protein involved in regulating Nif at the transcriptional level (50, 64). Thus, the transition of Nif from anoxic to oxic environments was associated with a shift from posttranslational regulation in anaerobes (NifH₁I₂) to transcriptional regulation (NifA) in obligate aerobes and facultative anaerobes. We hypothesize that the dynamic changes in the metabolic rate of aerobes due to fluctuating O₂ tensions in the environment and the higher rates of protein turnover that these conditions create would select for mechanisms that more efficiently regulate Nif at the transcriptional level. In the case of Nif, the selection pressure to increase the number of Nif-specific genes involved in optimizing the process of FeMo-co biosynthesis and Nif maturation and the transition from posttranslational to transcriptional regulation is likely associated with higher turnover rates and higher demands for fixed N because of the increased metabolic capacity associated with an aerobic lifestyle.

Conclusions. Evidence presented here is consistent with an origin for Nif in an anoxic environment and indicates a major increase in the number of *nif* genes during the transition of diazotrophs from anoxic to oxic ecological niches. Although it is not clear whether Nif emerged prior to the “Great Oxidation Event” (GOE) and the buildup of O₂ in the atmosphere or if it emerged after the GOE in an anoxic environment, it is clear that several innovations were key to enabling the adaptation of this O₂-sensitive process into oxic niches. A number of different protective mechanisms allow for functional Nif in aerobic bacteria, including temporal and spatial segregation and respiratory protection. Interestingly, our data show that the evolution of these protective mechanisms is not associated with increased complexity of *nif* gene contents. The majority of genes recruited during the transition from anoxic to oxic N₂ fixation are associated with metal cluster biosynthesis and regulation. Many of those implicated in metal cluster biosynthesis are not essential and are only required for maximal activity, indicating that they were recruited to refine Nif activity. In addition, our data show a clear transition from the NifH₁I₂-dependent posttranslational regulation of Nif activity in anaerobes to NifA-dependent transcriptional regulation in aerobes. The shift in regulation during the transition from anoxic to oxic niches suggests an underlying role for O₂ in driving this change in regulation. We hypothesize that both the increase in the number of *nif* genes and the shift in their mode of regulation during the transition of Nif from anaerobic to aerobic metabolism is due to selection for increased efficiency in the synthesis of Nif to meet the higher fixed N demands associated with more productive aerobic metabolism and to more efficiently regulate Nif under oxic conditions that favor protein turnover.

ACKNOWLEDGMENTS

This work was supported by grants from the NASA Astrobiology Institute (NNA15BB02A and NNA13AA94A) to E.S.B. and the National Science Foundation (IOS-1331098) to J.W.P.

REFERENCES

1. Falkowski PG. 1997. Evolution of the nitrogen cycle and its influence on the biological sequestration of CO₂ in the ocean. *Nature* 387:272–275. <http://dx.doi.org/10.1038/387272a0>.

2. Falkowski PG, Fenchel T, Delong EF. 2008. The microbial engines that drive Earth's biogeochemical cycles. *Science* 320:1034–1039. <http://dx.doi.org/10.1126/science.1153213>.
3. Boyd E, Peters JW. 2013. New insights into the evolutionary history of biological nitrogen fixation. *Front Microbiol* 4:201. <http://dx.doi.org/10.3389/fmicb.2013.00201>.
4. Peters JW, Boyd ES, Hamilton TL, Rubio LM. 2011. Biochemistry of Mo-nitrogenase, p 59–99. In Moir JWB (ed), *Nitrogen cycling in bacteria: molecular analysis*. Caister Academic Press, Norfolk, United Kingdom.
5. Boyd ES, Anbar AD, Miller S, Hamilton TL, Lavin M, Peters JW. 2011. A late methanogen origin for molybdenum-dependent nitrogenase. *Geobiology* 9:221–232. <http://dx.doi.org/10.1111/j.1472-4669.2011.00278.x>.
6. Boyd ES, Hamilton TL, Peters JW. 2011. An alternative path for the evolution of biological nitrogen fixation. *Front Microbiol* 2:205. <http://dx.doi.org/10.3389/fmicb.2011.00205>.
7. Raymond J, Siefert JL, Staples CR, Blankenship RE. 2004. The natural history of nitrogen fixation. *Mol Biol Evol* 21:541–554. <http://dx.doi.org/10.1093/molbev/msh047>.
8. Dos Santos P, Fang Z, Mason S, Setubal J, Dixon R. 2012. Distribution of nitrogen fixation and nitrogenase-like sequences amongst microbial genomes. *BMC Genomics* 13:162. <http://dx.doi.org/10.1186/1471-2164-13-162>.
9. Fay P. 1992. Oxygen relations of nitrogen fixation in cyanobacteria. *Microbiol Rev* 56:340–373.
10. Poole RK, Hill S. 1997. Respiratory protection of nitrogenase activity in *Azotobacter vinelandii*: roles of the terminal oxidases. *Biosci Rep* 17:303–317. <http://dx.doi.org/10.1023/A:1027336712748>.
11. Stal LJ, Krumbien WE. 1985. Nitrogenase activity in the non-heterocystous cyanobacterium *Oscillatoria* sp. grown under alternating light-dark cycles. *Arch Microbiol* 143:67–71. <http://dx.doi.org/10.1007/BF00414770>.
12. Dixon R, Eady RR, Espin G, Hill S, Iaccarino M, Kahn D, Merrick M. 1980. Analysis of regulation of *Klebsiella pneumoniae* nitrogen fixation (*nif*) gene cluster with gene fusions. *Nature* 286:128–132. <http://dx.doi.org/10.1038/286128a0>.
13. Rubio LM, Ludden PW. 2008. Biosynthesis of the iron-molybdenum cofactor of nitrogenase. *Annu Rev Microbiol* 62:93–111. <http://dx.doi.org/10.1146/annurev.micro.62.081307.162737>.
14. Enkh-Amgalan J, Kawasaki H, Seki T. 2006. Molecular evolution of the *nif* gene cluster carrying *nifH* and *nifD* genes in the Gram-positive phototrophic bacterium *Heliobacterium chlorum*. *Int J Syst Evol Microbiol* 56:65–74. <http://dx.doi.org/10.1099/ijs.0.63815-0>.
15. Dekas AE, Poretsky RS, Orphan VJ. 2009. Deep-sea Archaea fix and share nitrogen in methane-consuming microbial consortia. *Science* 326:422–426. <http://dx.doi.org/10.1126/science.1178223>.
16. Mehta MP, Baross JA. 2006. Nitrogen fixation at 92°C by a hydrothermal vent archaeon. *Science* 314:1783–1786. <http://dx.doi.org/10.1126/science.1134772>.
17. McGlynn SE, Boyd ES, Peters JW, Orphan VJ. 2012. Classifying the metal dependence of uncharacterized nitrogenases. *Front Microbiol* 3:419. <http://dx.doi.org/10.3389/fmicb.2012.00419>.
18. Arnold W, Rump A, Klipp W, Priefer UB, Pühler A. 1988. Nucleotide sequence of a 24,206-base-pair DNA fragment carrying the entire nitrogen fixation gene cluster of *Klebsiella pneumoniae*. *J Mol Biol* 203:715–738. [http://dx.doi.org/10.1016/0022-2836\(88\)90205-7](http://dx.doi.org/10.1016/0022-2836(88)90205-7).
19. Hamilton TL, Ludwig M, Dixon R, Boyd ES, Dos Santos PC, Setubal JC, Bryant DA, Dean DR, Peters JW. 2011. Transcriptional profiling of nitrogen fixation in *Azotobacter vinelandii*. *J Bacteriol* 193:4477–4486. <http://dx.doi.org/10.1128/JB.05099-11>.
20. Schmehl M, Jahn A, Meyer zu Vilsendorf A, Hennecke S, Masepohl B, Schuppler M, Marxer M, Oelze J, Klipp W. 1993. Identification of a new class of nitrogen fixation genes in *Rhodobacter capsulatus*: a putative membrane complex involved in electron transport to nitrogenase. *Mol Gen Genet* 241:602–615.
21. Herrmann G, Jayamani E, Mai G, Buckel W. 2008. Energy conservation via electron-transferring flavoprotein in anaerobic bacteria. *J Bacteriol* 190:784–791. <http://dx.doi.org/10.1128/JB.01422-07>.
22. Swofford DL. 2001. PAUP: phylogenetic analysis using parsimony (and other methods), version 4.0b10. Sinauer Associates, Sunderland, MA.
23. Abascal F, Zardoya R, Posada D. 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21:2104–2105. <http://dx.doi.org/10.1093/bioinformatics/bti263>.
24. Anisimova M, GM, Dufayard JF, Dessimoz C, Gascuel O. 2011. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Syst Biol* 60:685–699. <http://dx.doi.org/10.1093/sysbio/syr041>.
25. Sanderson MJ. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Mol Biol Evol* 19:101–109. <http://dx.doi.org/10.1093/oxfordjournals.molbev.a003974>.
26. Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289–290. <http://dx.doi.org/10.1093/bioinformatics/btg412>.
27. R Development Core Team. 2010. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
28. Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, Blomberg SP, Webb CO. 2010. Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* 26:1463–1464. <http://dx.doi.org/10.1093/bioinformatics/btq166>.
29. Blomberg SP, Garland JT, Ives AR. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* 57:717–745. <http://dx.doi.org/10.1111/j.0014-3820.2003.tb00285.x>.
30. Hu Y, Fay AW, Lee CC, Ribbe MW. 2007. P-cluster maturation on nitrogenase MoFe protein. *Proc Natl Acad Sci U S A* 104:10424–10429. <http://dx.doi.org/10.1073/pnas.0704297104>.
31. Hu Y, Fay AW, Dos Santos PC, Naderi F, Ribbe MW. 2004. Characterization of *Azotobacter vinelandii* *nifZ* deletion strains: indication of stepwise MoFe protein assembly. *J Biol Chem* 279:54963–54971. <http://dx.doi.org/10.1074/jbc.M408983200>.
32. Paul W, Merrick M. 1989. The roles of the *nifW*, *nifZ* and *nifM* genes of *Klebsiella pneumoniae* in nitrogenase biosynthesis. *Eur J Biochem* 178:675–682. <http://dx.doi.org/10.1111/j.1432-1033.1989.tb14497.x>.
33. Jacobson MR, Cash VL, Weiss MC, Laird NF, Newton WE, Dean DR. 1989. Biochemical and genetic analysis of the *nifUSVWZM* cluster from *Azotobacter vinelandii*. *Mol Gen Genet* 219:49–57.
34. Kim S, Burgess BK. 1994. Purification and characterization of nitrogenase from a $\Delta nifW$ strain of *Azotobacter vinelandii*. *J Biol Chem* 269:4216–4220.
35. Kim S, Burgess BK. 1996. Evidence for the direct interaction of the *nifW* gene product with the MoFe protein. *J Biol Chem* 271:9764–9770. <http://dx.doi.org/10.1074/jbc.271.16.9764>.
36. Hernandez JA, Curatti L, Aznar CP, Perova Z, Britt RD, Rubio LM. 2008. Metal trafficking for nitrogen fixation: NifQ donates molybdenum to NifEN/NifH for the biosynthesis of the nitrogenase FeMo-cofactor. *Proc Natl Acad Sci U S A* 105:11679–11684. <http://dx.doi.org/10.1073/pnas.0803576105>.
37. Imperial J, Ugalde RA, Shah VK, Brill WJ. 1984. Role of the *nifQ* gene product in the incorporation of molybdenum into nitrogenase in *Klebsiella pneumoniae*. *J Bacteriol* 158:187–194.
38. Rodríguez-Quinones F, Bosch R, Imperial J. 1993. Expression of the *nifBfdxNifOQ* region of *Azotobacter vinelandii* and its role in nitrogenase activity. *J Bacteriol* 175:2926–2935.
39. Jacobson MR, Brigle KE, Bennett LT, Setterquist RA, Wilson MS, Cash VL, Beynon J, Newton WE, Dean DR. 1989. Physical and genetic map of the major *nif* gene cluster from *Azotobacter vinelandii*. *J Bacteriol* 171:1017–1027.
40. Simon HM, Homer MJ, Roberts GP. 1996. Perturbation of *nifT* expression in *Klebsiella pneumoniae* has limited effect on nitrogen fixation. *J Bacteriol* 178:2975–2977.
41. Kessler PS, Daniel C, Leigh JA. 2001. Ammonia switch-off of nitrogen fixation in the methanogenic archaeon *Methanococcus maripaludis*: mechanistic features and requirement for the novel GlnB homologues, NifH1 and NifH2. *J Bacteriol* 183:882–889. <http://dx.doi.org/10.1128/JB.183.3.882-889.2001>.
42. Dodsworth JA, Leigh JA. 2006. Regulation of nitrogenase by 2-oxoglutarate-reversible, direct binding of a PII-like nitrogen sensor protein to dinitrogenase. *Proc Natl Acad Sci U S A* 103:9779–9784. <http://dx.doi.org/10.1073/pnas.0602278103>.
43. Dodsworth JA, Cady NC, Leigh JA. 2005. 2-Oxoglutarate and the PII homologues NifH1 and NifH2 regulate nitrogenase activity in cell extracts of *Methanococcus maripaludis*. *Mol Microbiol* 56:1527–1538. <http://dx.doi.org/10.1111/j.1365-2958.2005.04621.x>.
44. Bolhuis H, Severin I, Confurius-Guns V, Wollenzien UIA, Stal LJ. 2010. Horizontal transfer of the nitrogen fixation gene cluster in the cyanobacterium *Microcoleus chthonoplastes*. *ISME J* 4:121–130. <http://dx.doi.org/10.1038/ismej.2009.99>.

45. Rubio LM, Rangaraj P, Homer MJ, Roberts GP, Ludden PW. 2002. Cloning and mutational analysis of the gamma gene from *Azotobacter vinelandii* defines a new family of proteins capable of metallocluster binding and protein stabilization. *J Biol Chem* 277:14299–14305. <http://dx.doi.org/10.1074/jbc.M107289200>.
46. Krebs C, Agar JN, Smith AD, Frazzon J, Dean DR, Huynh BH, Johnson MK. 2001. IscA, an alternate scaffold for Fe-S cluster biosynthesis. *Biochemistry* 40:14069–14080. <http://dx.doi.org/10.1021/bi015656z>.
47. Leigh JA, Dodsworth JA. 2007. Nitrogen regulation in bacteria and archaea. *Annu Rev Microbiol* 61:349–377. <http://dx.doi.org/10.1146/annurev.micro.61.080706.093409>.
48. Gussin GN, Ronson CW, Ausubel FM. 1986. Regulation of nitrogen fixation genes. *Annu Rev Genet* 20:567–591. <http://dx.doi.org/10.1146/annurev.ge.20.120186.003031>.
49. Arcondeguy T, Jack R, Merrick M. 2001. P_{II} signal transduction proteins, pivotal players in microbial nitrogen control. *Microbiol Mol Biol Rev* 65:80–105. <http://dx.doi.org/10.1128/MMBR.65.1.80-105.2001>.
50. Dixon R, Kahn D. 2004. Genetic regulation of biological nitrogen fixation. *Nat Rev Microbiol* 2:621–631. <http://dx.doi.org/10.1038/nrmicro954>.
51. Martinez-Argudo I, Little R, Shearer N, Johnson P, Dixon R. 2004. The NifL-NifA system: a multidomain transcriptional regulatory complex that integrates environmental signals. *J Bacteriol* 186:601–610. <http://dx.doi.org/10.1128/JB.186.3.601-610.2004>.
52. Little R, Reyes-Ramirez F, Zhang Y, van Heeswijk WC, Dixon R. 2000. Signal transduction to the *Azotobacter vinelandii* NIFL-NIFA regulatory system is influenced directly by interaction with 2-oxoglutarate and the PII regulatory protein. *EMBO J* 19:6041–6050. <http://dx.doi.org/10.1093/emboj/19.22.6041>.
53. Hill S, Kennedy C, Kavanagh E, Goldberg RB, Hanau R. 1981. Nitrogen fixation gene (*nifH*) involved in oxygen regulation of nitrogenase synthesis in *K. pneumoniae*. *Nature* 290:424–426. <http://dx.doi.org/10.1038/290424a0>.
54. Ueki T, Lovley DR. 2010. Novel regulatory cascades controlling expression of nitrogen-fixation genes in *Geobacter sulfurreducens*. *Nucleic Acids Res* 38:7485–7499. <http://dx.doi.org/10.1093/nar/gkq652>.
55. Curatti L, Brown CS, Ludden PW, Rubio LM. 2005. Genes required for rapid expression of nitrogenase activity in *Azotobacter vinelandii*. *Proc Natl Acad Sci U S A* 102:6291–6296. <http://dx.doi.org/10.1073/pnas.0501216102>.
56. Sarkar A, Köhler J, Hurek T, Reinhold-Hurek B. 2012. A novel regulatory role of the Rnf complex of *Azoarcus* sp. strain BH72. *Mol Microbiol* 83:408–422. <http://dx.doi.org/10.1111/j.1365-2958.2011.07940.x>.
57. Edgren T, Nordlund S. 2004. The *fixABCX* genes in *Rhodospirillum rubrum* encode a putative membrane complex participating in electron transfer to nitrogenase. *J Bacteriol* 186:2052–2060. <http://dx.doi.org/10.1128/JB.186.7.2052-2060.2004>.
58. Benson DR, Silvester WB. 1993. Biology of *Frankia* strains, actinomycete symbionts of actinorhizal plants. *Microbiol Rev* 57:293–319.
59. Berry AM, Harriott OT, Moreau RA, Osman SF, Benson DR, Jones AD. 1993. Hopanoid lipids compose the *Frankia* vesicle envelope, presumptive barrier of oxygen diffusion to nitrogenase. *Proc Natl Acad Sci U S A* 90:6091–6094. <http://dx.doi.org/10.1073/pnas.90.13.6091>.
60. Kharbush JJ, Ugalde JA, Hogle SL, Allen EE, Aluwihare LI. 2013. Composite bacterial hopanoids and their microbial producers across oxygen gradients in the water column of the California Current. *Appl Environ Microbiol* 79:7491–7501. <http://dx.doi.org/10.1128/AEM.02367-13>.
61. Shah VK, Rangaraj P, Chatterjee R, Allen RM, Roll JT, Roberts GP, Ludden PW. 1999. Requirement of NifX and other *nif* proteins for in vitro biosynthesis of the iron-molybdenum cofactor of nitrogenase. *J Bacteriol* 181:2797–2801.
62. Hernandez JA, Igarashi RY, Soboh B, Curatti L, Dean DR, Ludden PW, Rubio LM. 2007. NifX and NifEN exchange NifB cofactor and the VK-cluster, a newly isolated intermediate of the iron-molybdenum cofactor biosynthetic pathway. *Mol Microbiol* 63:177–192. <http://dx.doi.org/10.1111/j.1365-2958.2006.05514.x>.
63. Moran NA. 2002. Microbial minimalism. *Cell* 108:583–586. [http://dx.doi.org/10.1016/S0092-8674\(02\)00665-7](http://dx.doi.org/10.1016/S0092-8674(02)00665-7).
64. Fischer HM. 1994. Genetic regulation of nitrogen fixation in rhizobia. *Microbiol Rev* 58:352–386.
65. Hill S, Kavanagh EP. 1980. Roles of *nifH* and *nifJ* gene products in electron transport to nitrogenase in *Klebsiella pneumoniae*. *J Bacteriol* 141:470–475.
66. Martin AE, Burgess BK, Iismaa SE, Smartt CT, Jacobson MR, Dean DR. 1989. Construction and characterization of an *Azotobacter vinelandii* strain with mutations in the genes encoding flavodoxin and ferredoxin I. *J Bacteriol* 171:3162–3167.
67. Bennett LT, Jacobson MR, Dean DR. 1988. Isolation, sequencing, and mutagenesis of the *nifF* gene encoding flavodoxin from *Azotobacter vinelandii*. *J Biol Chem* 263:1364–1369.
68. Roberts GP, MacNeil T, MacNeil D, Brill WJ. 1978. Regulation and characterization of protein products coded by the *nif* (nitrogen fixation) genes of *Klebsiella pneumoniae*. *J Bacteriol* 136:267–279.
69. Mapolelo DT, Zhang B, Naik SG, Huynh BH, Johnson MK. 2012. Spectroscopic and functional characterization of iron-sulfur cluster-bound forms of *Azotobacter vinelandii*^{Nif} IscA. *Biochemistry* 51:8071–8084. <http://dx.doi.org/10.1021/bi3006658>.
70. Howard KS, McLean PA, Hansen FB, Lemley PV, Koblan KS, Orme-Johnson WH. 1986. *Klebsiella pneumoniae* *nifM* gene product is required for stabilization and activation of nitrogenase iron protein in *Escherichia coli*. *J Biol Chem* 261:772–778.
71. Dixon R. 1998. The oxygen-responsive NIFL-NIFA complex: a novel two-component regulatory system controlling nitrogenase synthesis in γ -Proteobacteria. *Arch Microbiol* 169:371–380. <http://dx.doi.org/10.1007/s002030050585>.
72. Schmitz RA, Klopprogge K, Grabbe R. 2002. Regulation of nitrogen fixation in *Klebsiella pneumoniae* and *Azotobacter vinelandii*: NifL, transducing two environmental signals to the *nif* transcriptional activator NifA. *J Mol Microbiol Biotechnol* 4:235–242.
73. Homer MJ, Paustian TD, Shah VK, Roberts GP. 1993. The *nifY* product of *Klebsiella pneumoniae* is associated with apodinitrogenase and dissociates upon activation with the iron-molybdenum cofactor. *J Bacteriol* 175:4907–4910.
74. Rubio LM, Singer SW, Ludden PW. 2004. Purification and characterization of NafY (apodinitrogenase γ subunit) from *Azotobacter vinelandii*. *J Biol Chem* 279:19739–19746. <http://dx.doi.org/10.1074/jbc.M400965200>.
75. Zheng L, White RH, Dean DR. 1997. Purification of the *Azotobacter vinelandii* *nifV*-encoded homocitrate synthase. *J Bacteriol* 179:5963–5966.
76. Zheng L, White RH, Cash VL, Jack RF, Dean DR. 1993. Cysteine desulfurase activity indicates a role for NIFS in metallocluster biosynthesis. *Proc Natl Acad Sci U S A* 90:2754–2758. <http://dx.doi.org/10.1073/pnas.90.7.2754>.